DOCUMENT RESUME

ED 247 409 CE 039 438

AUTHOR	Misanchuk, Earl R.
TITLE	The Analysis of Multi-Component Educational and Training Needs Data: The Multivariate Case.
PUB DATE	Jan 84
NOTE	35p.; Paper presented at the Annual Meeting of the Association of Educational Communications and Technology (Dallas, TX, January 20-24, 1984).
PUB TYPE	Viewpoints (120) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Comparative Analysis; Competence; *Data Analysis; *Educational Needs; Educational Research; Models; *Multivariate Analysis; *Needs Assessment; Relevance (Education); Research Design; *Research Methodology; Student Educational Objectives

ABSTRACT

The technique of multivariate analysis is particularly suited to educational needs assessment research because it allows for the summarization of data across any number of learners or components of educational need to produce a single numerical index of need for each skill examined. In the needs assessment process, educational or training need is assumed to have three underlying dimensions: competence of the individual at a task or skill, relevance of the task or skill to the individual, and the individual's desire to further his or her learning of the task or skill. The use of multivariate analysis as a means for assessing educational need is superior to previous methods in that these earlier models could accommodate data for only two dimensions. In addition, the multivariate analysis method manifests an increased sensitivity to changes in respondent distribution in the index of educational need. Finally, the model can accommodate different numbers and relative emphases of dimensions according to user-defined models and its method of computation is a relatively simple one. (This paper includes a discussion of the multivariate assessment technique as a model for assessing educational and training need, discussions of bivariate and multivariate cases of analysis, a computational example, a list of references, and six figures and one table illustrating various phases of the model and its standard error weights.) (MN)

THE ANALYSIS OF MULTI-COMPONENT EDUCATIONAL AND TRAINING NEEDS DATA: THE MULTIVARIATE CASE

Earl R. Misanchuk[1]
Division of Extension and Community Relations
University of Saskatchewan
Saskatoon, Saskatchewan S7N 0W0

Paper presented at the Annual Meeting of the Association of Educational Communications and Technology, Dallas, TX, January 20-24, 1984

Abstract

Instructional developers are becoming increasingly involved in needs assessment, determining what needs to be taught as well as determining how best to teach and evaluate it. As increasingly complex models of the needs assessment process are proposed, methods must be developed to implement them. A three-component model of educational and training needs requires examining data on the competence of an individual at a skill, the relevance to the individual of the skill, and the individual's desire for further learning on the skill. Such data collections can be mind-boggling to analyze, especially if there are a significant number of skills and/or learners involved. This paper describes and illustrates a statistic which can summarize data across any number of learners or components of educational need, to produce a single numerical index of need for each skill.

As instructional developers increasingly become involved with the business of finding out what needs to be taught, as well as determining how best to teach and evaluate it, attention is focussed on the processes of needs identification and analysis. Whether in educational institutions or in training environments, instructional developers are being exposed to a growing body of literature reflecting increasing awareness of and attention to this not-yet-all-that-clearly-understood technology (see, for example, Deden-Parker, 1980; Gagne, 1977; Jones & Sommers, 1975; Kaufman, 1972, 1977a, 1977b; Kaufman & English, 1979; Kaufman, Stakenas, Wager, & Mayer, 1981; Misanchuk, in press; Monette, 1977; Rossett, 1982; Roth, 1977; Scissons, 1982; Scriven & Roth, 1977; Spitzer, 1981; Witkin, 1976, 1977).

The emergence of formalized models of needs assessment (e.g., Kaufman, 1977a, 1977b; Kaufman & English, 1979; Witkin, 1976, 1977) indicates the growing maturity of the field. As new models are explored, techniques for implementing them must be developed. For example, multidimensional conceptualizations of educational and training need have been put forward (Misanchuk, 1982; Misanchuk & Scissons, 1978; Scissons, 1982), but because of the difficulty in extracting comprehensible information from multidimensional data, have not been widely applied.

We can illustrate the problem by focussing on the model wherein educational or training need[2] is assumed to have three underlying dimensions: competence of the individual at the task/skill[3], relevance to the individual of the task/skill, and the individual's desire to further his or her learning of the task/skill (Misanchuk, 1982). High educational need is defined, in terms of these dimensions, when an individual demonstrates low competence, high relevance, and high desire. Similarly, a particular skill can be said to have high need associated with it when a pertinent group of individuals demonstrates --on the average--low competence, high relevance, and high desire.

It is straight-forward enough, albeit not trivial, to collect information about each individual's competence at a task, the relevance of that task to the individual, and the individual's desire for further training at the task. For example, five-point Likert-type scales could be used to collect data on each of the three dimensions. Once collected, however, the many data points--one for

each of the three dimensions, for each training task for which need is being assessed, for each individual in the learner group; a total of 3*ts* data points, where *t* = the number of skills whose needs are being assessed, and *s* = the number of potential learners involved--prove resistant to the analyst's traditional efforts. Several approaches to such analyses have been attempted and found wanting, including arbitrarily-defined cutting points with *z*-scores computed using the pooled inter-item variance (Misanchuk & Scissons, 1978), the odds ratio (Reynolds, 1977), and the dichotomized-additive coefficient procedure (LeSage, 1980; Misanchuk, 1980). If it were possible to reduce the 3*ts* data points to, say, *t* data points, so that each skill would have associated with it an index of educational need which could be compared to the corresponding indexes for other skills, the results of the needs analysis would be much more comprehensible.

A Multivariate Model of Educational Need

For simplicity in language, the discussion here will focus on a three-dimensional model of educational need. However, the argument can be extended to *n* dimensions.

Consider the cube in Figure 1, respresenting the postulated three-dimensional configuration (Misanchuk, 1982) of an educational need. According to the model, the stippled cell in the forward-most, upper-right corner represents the position of highest need. In other words, if all individuals being surveyed responded to each of the three dimensions being measured in such a way as to place themselves in that forward, upper-right cell, the skill about which the individuals were being asked would have associated with it an extremely high educational need.

Insert Figure 1 about here.

Conversely, if all individuals were in the rear-most, lower left cell (which is completely hidden from view in Figure 1), the skill under consideration would be said to have virtually no educational need associated

with it.

Problems of interpretation occur when, as is usual in real life, the data distribute themselves across all three dimensions. Some examples will illustrate these difficulties, but first a diversion is necessary to establish a convention which will make it easier to describe individual cells in an array such as Figure 1.

For convenience in communication, the individual cells in the three-dimensional matrix will be designated in terms of their row, column, and layer (or stratum) numbers. Thus, (1,5,1) represents the cell in row 1, column 5, and layer 1 (the forward-most). Cell (1,5,1), is of course, the cell indicating highest need. The cell associated with lowest need, mentioned earlier, is (5,1,5), and the cross-hatched cell in Figure 1 is (2,4,1), while the cell with horizontal shading is (4,5,2). Cell (3,3,3) is in the exact center of the cube.

Now the examples: It should be intuitively obvious that a skill for which all responses fall into cell (3,3,3) has a lower educational need than does one for which all responses fall into cell (4,2,4), but how does it compare to a skill for which all responses fall into cell (3,2,2)? What if, for one skill, there are 17 people in (3,2,2), while for another skill, there are 15 in cell (3,3,3): which skill demonstrates the higher need? In real life, we can expect to find varying numbers of individuals in each of the 125 cells in Figure 1, further complicating matters.

One solution to the problem wherein only two dimensions were incorporated into the model of educational need (Misanchuk, in press) adapted a proportionate reduction in error statistic, del (∇) (Hildebrand, Laing, and Rosenthal, 1977a, 1977b), for use in needs analysis. The ensuing statistic, $\nabla_{\underline{N}}$ (the proportionate reduction in error index of educational need), can collapse information from two dimensions into a single statistic, which represents the educational need associated with the skill under consideration. For example, educational need can be defined in terms of both relevance and competence, and a single value can be used to represent the average measure on both dimensions simultaneously. Unfortunately, the symmetry of the mathematics involved in the

analysis permits only a two-dimensional model of need.

The remainder of this paper describes an extension of V_N, which can be applied to n-dimensional situations. For the sake of simplicity and clarity, the development of the procedure will use a three-dimensional model, but the extension is straightforward.

The underlying logic of the PRE approach involves predicting the probability that certain combinations of a joint distribution will occur, then testing to see how closely the prediction matches observations. In the current application, we postulate a high educational need for a given skill, then compare the observed distribution of responses given by learners to the prediction. In more concrete terms, we begin by predicting that all respondents will answer in cell (1,5,1), then apply a mathematical procedure based on a proportionate reduction in error approach (Hildebrand, et al., 1977a, 1977b) to determine how accurate our prediction was. The result of the procedure, the statistic V_N, can be used as an index of the educational need associated with the skill under consideration, provided certain assumptions are made and their implications incorporated into the calculation (Misanchuk, in press).

The Bivariate Case

In the two-component case, using competence and relevance for the sake of illustrating the procedure, high need would be defined as concomitant low competence and high relevance. Our prediction of high need therefore translates into a prediction of low competence and high relevance. In graphic terms, we would be dealing with only the front-most layer of the cube in Figure 1 in the two-component case; high need is indicated by cell (1,5,1). Any learners' responses falling in other cells would constitute errors in our prediction. Obviously, the further the cells in question are from (1,5,1) in both directions, the more severe the effects of the errors on the accuracy of the prediction.

The PRE procedure permits the assignment of varying "degrees of error" to

each cell to accommodate the increased severity of effect. Thus cell (1,5,1) would have no error associated with the prediction (i.e., an error weight of 0), while cells (1,4,1) and (2,5,1) could be given error weights of 0.177, cell (2,4,1) could be given an error weight of 0.250, and so forth. The "worst" error, cell (5,1,1), could be considered a "whole" error, and have an error weight of 1.0 . While the numerical values of the error weights are arbitrary—others could be assigned at the discretion of the researcher—they are not chosen capriciously: arguments have been presented for the values suggested here (Misanchuk, in press).

The statistic del is defined by Hildebrand et al. (1977a, 1977b) as

$$\nabla = \frac{\text{Expected errors} - \text{Observed errors}}{\text{Expected errors}},$$

with both expected and observed errors taking into account the cell error weights mentioned above. The expected error rate is determined by the marginal totals of the matrix in the same way as in chi square.[4]

In formal terms, del is defined as

$$\nabla = 1 - \frac{\sum_{\underline{i}=1}^{\underline{R}} \sum_{\underline{j}=1}^{\underline{C}} \underline{W}_{\underline{ij}}\, \underline{P}_{\underline{ij}}}{\sum_{\underline{i}=1}^{\underline{R}} \sum_{\underline{j}=1}^{\underline{C}} \underline{W}_{\underline{ij}}\, \underline{P}_{\underline{i}.}\, \underline{P}_{.\underline{j}}}, \qquad (1)$$

where $\underline{W}_{\underline{ij}}$ is the error weight for cell ($\underline{i},\underline{j}$) ($\underline{W}_{\underline{ij}} = 1$ for every "whole" error cell; $0 < \underline{W}_{\underline{ij}} < 1$ for every "partial" error cell), $\underline{P}_{\underline{ij}}$ is the probability of a randomly sampled observation falling into cell ($\underline{i},\underline{j}$), and $\underline{P}_{\underline{i}.}$ and $\underline{P}_{.\underline{j}}$ are the expected marginal probabilities for the rows and columns, respectively.[5] The

proportionate reduction in error index of educational need, ∇_{N} (Misanchuk, in press), is computationally equivalent to Formula (1), but assumes certain known values for the error weights and the expected marginal probabilities.

For the bivariate case, it matters not which of the two variables is designated the dependent variable and which is designated the independent; both ways, the computed value of ∇ is the same (Hildebrand, et al., 1977b, p. 71). (Indeed, in the needs analysis situation, the designation of variables as dependent and independent is quite meaningless.) Unfortunately, this symmetry does not hold for the multivariate case, and despite the lack of meaning associated with the terms dependent and independent for needs analysis, some accommodation must be made for that fact.

The Multivariate Case

Hildebrand, et al. (1977a, 1977b) show that the PRE approach can be used with $\underline{n}$-dimensional arrays by collapsing them into two-dimensional ones: the independent variables are reconfigured as a single (new) independent variable composed of all the possible combinations of the original independent variables.

To illustrate, assume a dependent variable $\underline{Y}$ with levels $\underline{y}_1$, $\underline{y}_2$, $\underline{y}_3$, and two independent variables $\underline{X}$ and $\underline{Z}$ with levels $\underline{x}_1$ and $\underline{x}_2$, and $\underline{z}_1$, $\underline{z}_2$, and $\underline{z}_3$ respectively. Graphically, this example arrangement would be that displayed in Figure 2(a). The strategy for collapsing is simple: just change the labels on the columns to reflect the conjunction of the two variables. Instead of having two columns labelled $\underline{z}_1$—one under $\underline{x}_1$ and one under $\underline{x}_2$—we label the two columns $\underline{x}_1$ & $\underline{z}_1$, and $\underline{x}_2$ & $\underline{z}_1$.

- - - - - - - - - - - - - - - - -

Insert Figure 2 about here.

- - - - - - - - - - - - - - - - -

Thus, the collapsing process advocated by Hildebrand et al. (1977a, 1977b) creates a new variable $\underline{V}$ whose elements are $\underline{x}_1$ & $\underline{z}_1$, $\underline{x}_1$ & $\underline{z}_2$, $\underline{x}_1$ & $\underline{z}_3$, $\underline{x}_2$ & $\underline{z}_1$,

$\underline{x}_2$ & $\underline{z}_2$, and $\underline{x}_2$ & $\underline{z}_3$, as shown in Figure 2(b). The computation of ∇ then proceeds in the same manner as in the bivariate case, using $\underline{Y}$ as the dependent variable (with three levels) and $\underline{V}$ as the independent variable (with six levels).

Formally, Hildebrand et al. (1977b, p. 261) define the trivariate ∇ as

$$\nabla_{\underline{YXZ}} = 1 - \frac{\sum_{\underline{i}=1}^{\underline{R}} \sum_{\underline{j}=1}^{\underline{C}} \sum_{\underline{k}=1}^{\underline{S}} \underline{W}_{\underline{ijk}} \underline{P}_{\underline{ijk}}}{\sum_{\underline{i}=1}^{\underline{R}} \sum_{\underline{j}=1}^{\underline{C}} \sum_{\underline{k}=1}^{\underline{S}} \underline{W}_{\underline{ijk}} \underline{P}_{\underline{i}..} \underline{P}_{.\underline{jk}}} \tag{2}$$

Formula (2) applies when $\underline{Y}$ is the dependent variable (with $\underline{i}$ levels) and $\underline{X}$ and $\underline{Z}$ are the independent variables (with $\underline{j}$ and $\underline{k}$ levels, respectively). If either $\underline{X}$ or $\underline{Z}$ is chosen as the dependent variable, some of the dot notation in the formula must change to reflect the change in dependent variable. Specifically, the dot notation in the denominator of Formula (2) is $\underline{P}_{.\underline{j}.} \underline{P}_{\underline{i}.\underline{k}}$ when $\underline{X}$ is the dependent variable, and $\underline{P}_{..\underline{k}} \underline{P}_{\underline{ij}.}$ when $\underline{Z}$ is the dependent variable.[6] The different formulas thus generated yield different values for ∇, depending on which of the three variables is designated as the dependent variable, hence the symmetry problem alluded to earlier.

By analogy to the bivariate case, the approach to the computation of the multivariate PRE index of educational need considers that cell (1,5,1) is errorless (i.e., its error weight $\underline{W} = 0$), and that all other cells have associated error weights which increase as one moves away from cell (1,5,1) in any one or some combination of the $\underline{n}$ dimensions. For the trivariate case, cell (5,1,5) is assumed to have an error weight of 1.0, and all others have proportionate weights. The weights in Figure 3 are based on a simple geometric proportioning through three dimensions of the distance between cells (1,5,1) and (5,1,5), and are presented here as the proposed standard distribution of error weights for trivariate needs analysis using five-point scales.

Obviously, other distributions of weights could be used if there seems a need to do so.

- - - - - - - - - - - - - - - - -

Insert Figure 3 about here.

- - - - - - - - - - - - - - - - -

To handle the symmetry problem, we propose that, since the notion of dependent and independent variables is in any case irrelevant to the needs analysis situation, the n variables involved in the analysis (typically three: relevance, competence, and desire) each be treated in turn as the dependent variable, and the resulting values of ∇ be averaged.

Computational Example

To illustrate the procedure, an intuitive approach will be taken to a step-by-step calculation of a three-dimensional generic ∇, which will subsequently be contrasted with the calculation of a three-dimensional PRE index of educational need, $\nabla_{\underline{N}_3}$. This will be followed by the expression of a computational formula for the n-dimensional PRE index of educational need, $\nabla_{\underline{N}_n}$, and by several examples designed to provide some understanding of the range of values of $\nabla_{\underline{N}_3}$ that can be expected for different distributions of data.

Consider the array of data displayed in Figure 4(a). The dependent variable, competence, is displayed as the row variable, and the columns are formed by all possible combinations of levels of relevance and desire. The marginal totals are computed by simply summing across the rows and columns. Entries of 0 have been eliminated both within cells and in marginals to avoid visual clutter.

- - - - - - - - - - - - - - - - -

Insert Figure 4 about here.

- - - - - - - - - - - - - - - - -

By substitution into Formula (2), working by columns and substituting only for non-empty cells (since, for empty cells, $\underline{P}_{ijk} = 0$), and using error weights from Figure 3, we have as the numerator of the second term

$$\begin{aligned}&(.1443)\ 3/63 + (.2041)\ 9/63 + (.4564)\ 2/63 + (.0000)\ 7/63 + (.1443)\ 11/63\ + \\ &(.4787)\ 1/63 + (.2041)\ 2/63 + (.2500)\ 6/63 + (.1443)\ 3/63 + (.2041)\ 1/63\ + \\ &(.3227)\ 4/63 + (.4330)\ 2/63 + (.3227)\ 4/63 + (.6770)\ 3/63 + (.6292)\ 1/63\ + \\ &(.4564)\ 1/63 + (.8416)\ 2/63 + (.6124)\ 1/63\ .\end{aligned} \tag{3}$$

In the denominator of the second term, the row marginals, $\underline{P}_{i..}$, and the column marginals, $\underline{P}_{.jk}$, are multiplied by the appropriate cell error weight $\underline{W}_{ijk}$ for every cell where both $\underline{P}_{i..}$ and $\underline{P}_{.jk}$ are non-zero:

$$\begin{aligned}&(.1443)(15/63)(14/63) + (.2041)(36/63)(14/63) + (.3227)\ (7/63)(14/63)\ + \\ &(.4564)\ (5/63)(14/63) + (.0000)(15/63)(18/63) + (.1443)(36/63)(18/63)\ + \\ &(.2887)\ (7/63)(18/63) + (.4330)\ (5/63)(18/63) + (.4564)(15/63)\ (1/63)\ + \\ &(.4787)(36/63)\ (1/63) + (.5401)\ (7/63)\ (1/63) + (.6292)\ (5/63)\ (1/63)\ + \\ &(.2041)(15/63)\ (8/63) + (.2500)(36/63)\ (8/63) + (.3536)\ (7/63)\ (8/63)\ + \\ &(.4787)\ (5/63)\ (8/63) + (.1443)(15/63)\ (8/63) + (.2041)(36/63)\ (8/63)\ + \\ &(.3227)\ (7/63)\ (8/63) + (.4564)\ (5/63)\ (8/63) + (.4082)(15/63)\ (2/63)\ + \\ &(.4330)(36/63)\ (2/63) + (.5000)\ (7/63)\ (2/63) + (.5951)\ (5/63)\ (2/63)\ + \\ &(.2887)(15/63)\ (4/63) + (.3227)(36/63)\ (4/63) + (.4082)\ (7/63)\ (4/63)\ + \\ &(.5204)\ (5/63)\ (4/63) + (.6124)(15/63)\ (3/63) + (.6292)(36/63)\ (3/63)\ + \\ &(.6770)\ (7/63)\ (3/63) + (.7500)\ (5/63)\ (3/63) + (.4564)(15/63)\ (1/63)\ + \\ &(.4787)(36/63)\ (1/63) + (.5401)\ (7/63)\ (1/63) + (.6292)\ (5/63)\ (1/63)\ + \\ &(.4330)(15/63)\ (1/63) + (.4564)(36/63)\ (1/63) + (.5204)\ (7/63)\ (1/63)\ + \\ &(.6214)\ (5/63)\ (1/63) + (.7217)(15/63)\ (2/63)\ \ (.7360)(36/63)\ (2/63)\ + \\ &(.7773)\ (7/63)\ (2/63) + (.8416)\ (5/63)\ (2/63) + (.5951)(15/63)\ (1/63)\ + \\ &(.6124)(36/63)\ (1/63) + (.6614)\ (7/63)\ (1/63) + (.7360)\ (5/63)\ (1/63)\ .\end{aligned} \tag{4}$$

Simplifying,

$$V = 1 - \frac{0.2643}{0.2775} = 0.0476 \text{ with competence as the dependent variable.}$$

Recasting the same data so that relevance is the dependent (row) variable

and competence and desire are the independent (column) variables, we get the matrix found in Figure 4(b). Because the error weights migrate with the cells—they are properties of the cells, determined by the relative values of the variables that define the cells, rather than properties of the order in which the cells are arranged for display—the error weights substituted into Formula (2) are now those in Figure 5(a). The numerator of the second term becomes

$$(.1443)\ 3/63 + (.0000)\ 7/63 + (.2041)\ 9/63 + (.1443)\ 11/63 + (.4564)\ 2/63 + (.2041)\ 2/63 + (.1443)\ 3/63 + (.4787)\ 1/63 + (.2500)\ 6/63 + (.2041)\ 1/63 + (.3227)\ 4/63 + (.4330)\ 2/63 + (.3227)\ 4/63 + (.4564)\ 1/63 + (.6770)\ 3/63 + (.6292)\ 1/63 + (.6124)\ 1/63 + (.8416)\ 2/63 \tag{5}$$

$$= 0.2643,$$

and the denominator of that term becomes

$$\begin{aligned}
&(.4330)(6/63)(10/63) + (.2887)(2/63)(10/63) + (.1443)(24/63)(10/63) + \\
&(.0000)(31/63)(10/63) + (.4564)(6/63)(20/63) + (.3227)(2/63)(20/63) + \\
&(.2041)(24/63)(20/63) + (.1443)(31/63)(20/63) + (.6124)(6/63)(2/63) + \\
&(.5204)(2/63)(2/63) + (.4564)(24/63)(2/63) + (.4330)(31/63)(2/63) + \\
&(.4564)(6/63)(5/63) + (.3227)(2/63)(5/63) + (.2041)(24/63)(5/63) + \\
&(.1443)(31/63)(5/63) + (.4787)(6/63)(8/63) + (.3536)(2/63)(8/63) + \\
&(.2500)(24/63)(8/63) + (.2041)(31/63)(8/63) + (.5401)(6/63)(4/63) + \\
&(.4330)(2/63)(4/63) + (.3536)(24/63)(4/63) + (.3227)(31/63)(4/63) + \\
&(.5401)(6/63)(6/63) + (.4330)(2/63)(6/63) + (.3536)(24/63)(6/63) + \\
&(.3227)(31/63)(6/63) + (.6292)(6/63)(1/63) + (.5401)(2/63)(1/63) + \\
&(.4787)(24/63)(1/63) + (.4564)(31/63)(1/63) + (.6770)(6/63)(3/63) + \\
&(.5951)(2/63)(3/63) + (.5401)(24/63)(3/63) + (.5204)(31/63)(3/63) + \\
&(.7500)(6/63)(1/63) + (.6770)(2/63)(1/63) + (.6292)(24/63)(1/63) + \\
&(.6124)(31/63)(1/63) + (.7360)(6/63)(1/63) + (.6614)(2/63)(1/63) + \\
&(.6124)(24/63)(1/63) + (.5951)(31/63)(1/63) + (.8416)(6/63)(2/63) + \\
&(.7773)(2/63)(2/63) + (.7360)(24/63)(2/63) + (.7217)(31/63)(2/63)
\end{aligned} \tag{6}$$

$$= 0.2777 .$$

Notice that, as one might expect from perusal of Formula (2), the numerator has the same terms as it did when competence was the dependent variable, although they appear in a different order due to the reconfiguration of the data.

Insert Figure 5 about here.

With relevance as the dependent variable, then,

$$\nabla = 1 - \frac{0.2643}{0.2777} = 0.0483 .$$

In a similar manner, the data can be reconfigured again, as in Figure 4(c), making desire the (dependent) row variable, and relevance and competence the (independent) column variables. Substituting these reconfigured values into Formula (2), and using the error weights as now arranged in Figure 5(b), yields a value of $\nabla = 0.0557$.

So far in this example, the marginal totals have been substituted directly into the computational formula. However, as in the bivariate case, it makes less sense to use the observed marginals than it does to make some assumptions about the marginals (see Misanchuk, in press).

By analogy to the bivariate case, both competence and relevance will be assumed to monotonically increase along both dimensions as we move away from cell (1,1,1). The rationale for this assumption is that the task analysis underlying the needs assessment process should have ensured that the tasks used as a basis for the needs identification procedure were, by and large, quite relevant to the job roles of the individuals being assessed, and that the process of natural selection that obtains in hiring and firing tends to place more-or-less competent individuals into job roles. Therefore expected probability distributions of 0, 0.1, 0.2, 0.3, and 0.4 are probably reasonable (Misanchuk, in press).

With respect to desire, however, there is no justification to assume a monotonically increasing distribution. It does seem reasonable to assume a normal distribution, on the grounds that in any population, one could expect to find varying degrees of enthusiasm for undertaking further training on any particular skill, regardless of relevance of the skill or competence in it.[7] Superimposing the unit normal curve on a five-point scale gives the expected marginal distribution for desire of 0.0359, 0.2384, 0.4514, 0.2384, and 0.0359, which is proposed here as the standard expected marginal distribution for desire to undertake further training, and which will be used in the remainder of this paper. Again, if the needs analyst has information about the population that makes this assumption unwarranted, that information could be translated into a more acceptable expected marginal distribution for desire.

With these assumptions about marginal distributions, substitution into Formula (2) is changed somewhat, since instead of the observed marginals, we must use the assumed marginals. For Figure 4(a), the row marginals are therefore 0, 0.1, 0.2, 0.3, 0.4 . The column marginals are formed by the multiplication of the expected marginals for the two variables making up each of the columns, namely, 0, 0.1, 0.2, 0.3, 0.4 for relevance, and 0.0359, 0.2384, 0.4514, 0.2384, 0.0359 for desire. Thus 0, 0.0036, 0.0072, 0.0108, 0.0144, 0, 0.0238, 0.0476, 0.0714, 0.0952, 0, 0.0451, 0.0902, 0.1353, 0.1804, 0, 0.0238, 0.0476, 0.0714, 0.0952, 0, 0.0036, 0.0072, 0.0108, and 0.0144 form the column marginals for Figure 4(a).

With these marginals and the error weights in Figure 3, the denominator of Formula (2) for the data in Figure 4(a) becomes

$$
\begin{array}{llll}
(.5774)\ (0)\ (0) & +\ (.5951)\ (.1)\ (0) & +\ (.6455)\ (.2)\ (0) & + \\
(.7217)\ (.3)\ (0) & +\ (.8165)\ (.4)\ (0) & +\ (.4330)\ (0)\ (.0036) & + \\
(.4564)\ (.1)\ (.0036) & +\ (.5204)\ (.2)\ (.0036) & +\ (.6124)\ (.3)\ (.0036) & + \\
(.7271)\ (.4)\ (.0036) & +\ (.2887)\ (0)\ (.0072) & +\ (.3227)\ (.1)\ (.0072) & + \\
 & \vdots & & \\
(.6455)\ (.2)\ (.0144) & +\ (.7217)\ (.3)\ (.0144) & +\ (.8165)\ (.4)\ (.0144)\ . &
\end{array}
\qquad (7)
$$

Note that the numerator of Formula (2) is unaffected by the assumption of expected marginal distributions, and remains 0.2643 . Substituting Term (7) into Formula (2) and simplifying, we get

$$\nabla_{\underline{N}} = 1 - \frac{0.2643}{0.5748} = 0.5402 \text{ .}$$

Similar substitutions of expected marginals—arranged in appropriate order for the data displayed in Figures 4(b) and 4(c)—and error weights from Figures 5(a) and 5(b), respectively, give values of $\nabla_{\underline{N}}$ of 0.5402 and 0.5414 .

The mean of the three values of $\nabla_{\underline{N}}$ is considered the trivariate proportionate reduction in error index of educational need, $\nabla_{\underline{N}_3}$, and is equal in this case to 0.5406 .

In more formal terms, the trivariate PRE index of educational need, $\nabla_{\underline{N}_3}$, is

$$\nabla_{\underline{N}_3} = 1 - \frac{1}{3}\left[\frac{\sum_{\underline{i}=1}^{\underline{R}}\sum_{\underline{j}=1}^{\underline{C}}\sum_{\underline{k}=1}^{\underline{S}} \underline{w}_{\underline{ijk}}\,\underline{p}_{\underline{ijk}}}{\sum_{\underline{i}=1}^{\underline{R}}\sum_{\underline{j}=1}^{\underline{C}}\sum_{\underline{k}=1}^{\underline{S}} \underline{w}_{\underline{ijk}}\,\underline{p}_{\underline{i}..}\,\underline{p}_{.\underline{jk}}} + \frac{\sum_{\underline{i}=1}^{\underline{R}}\sum_{\underline{j}=1}^{\underline{C}}\sum_{\underline{k}=1}^{\underline{S}} \underline{w}_{\underline{ijk}}\,\underline{p}_{\underline{ijk}}}{\sum_{\underline{i}=1}^{\underline{R}}\sum_{\underline{j}=1}^{\underline{C}}\sum_{\underline{k}=1}^{\underline{S}} \underline{w}_{\underline{ijk}}\,\underline{p}_{.\underline{j}.}\,\underline{p}_{\underline{i}.\underline{k}}} + \frac{\sum_{\underline{i}=1}^{\underline{R}}\sum_{\underline{j}=1}^{\underline{C}}\sum_{\underline{k}=1}^{\underline{S}} \underline{w}_{\underline{ijk}}\,\underline{p}_{\underline{ijk}}}{\sum_{\underline{i}=1}^{\underline{R}}\sum_{\underline{j}=1}^{\underline{C}}\sum_{\underline{k}=1}^{\underline{S}} \underline{w}_{\underline{ijk}}\,\underline{p}_{..\underline{k}}\,\underline{p}_{\underline{ij}.}}\right] \quad (8)$$

given some pre-specified expected marginal distributions and error weights.

Notice that Formula (8) is composed of the three different versions—in terms of dotted subscripts—of the $\nabla_{\underline{N}}$ version of Formula (2), which are then averaged.

The extension to n dimensions is reasonably easy, if notationally complex. To avoid the unnecessary complication of extending the notation into a general formula for the n-dimensional case, we will simply describe the extension conceptually: The fraction 1/3 becomes 1/n, where n = the number of dimensions; there should be n terms inside the square brackets, each with the

same numerator shown in Formula (8), except that the subscripts for both $\underline{W}$ and $\underline{P}$ should number $\underline{n}$ (i.e., the subscripts should be $\underline{W}_{ijkm}$ and $\underline{P}_{ijkm}$ for four dimensions, $\underline{W}_{ijkmn}$ and $\underline{P}_{ijkmn}$ for five dimensions, etc.), and there should be $\underline{n}$ summation signs. Also, the denominators of the $\underline{n}$ terms within the brackets should have the dot notation

$$\underline{P}_{i\ldots}\ \underline{P}_{.jkm},\ \underline{P}_{.j..}\ \underline{P}_{i.km},\ \underline{P}_{..k.}\ \underline{P}_{ij.m},\ \underline{P}_{...m},\ \underline{P}_{ijk.},$$

respectively, for the four-dimensional case;

$$\underline{P}_{i....}\ \underline{P}_{.jkmn},\ \underline{P}_{.j...}\ \underline{P}_{i.kmn},\ \underline{P}_{..k..}\ \underline{P}_{ij.mn},\ \underline{P}_{...m.}\ \underline{P}_{ijk.n},\ \underline{P}_{....n}\ \underline{P}_{ijkm.}$$

for the five-dimensional case; and so forth. Again, there should be $\underline{n}$ summation signs and $\underline{n}$ subscripts for $\underline{W}$.

Some Illustrative Examples of $V_{\underline{N}_3}$

To provide a feeling for the kinds of values one can expect for $V_{\underline{N}_3}$, the data in Figure 6, along with the error weights in Figure 3 and the expected marginal distributions discussed earlier, have been substituted into Formula (8). The data in the various matrixes in Figure 6 have been arbitrarily arranged in a way that keeps the total number of observations constant, but systematically moves increasing numbers of responses away from the cell associated with highest need. Table 1 shows, as might be expected, that the values of $V_{\underline{N}_3}$ decrease as we move from the data arrangement in Figure 6(a) to that in Figure 6(f).

- - - - - - - - - - - - - - - - -

Insert Figure 6 about here.

- - - - - - - - - - - - - - - - -

- - - - - - - - - - - - - - - - -

Insert Table 1 about here.

- - - - - - - - - - - - - - - - -

Conclusion

The proportionate reduction in error (PRE) index of educational need, V_N (Misanchuk, in press), which could accommodate data from only two dimensions, has been defined in a more general way, permitting n-dimensional data to be analyzed. The expanded definition—the multivariate proportionate reduction in error index of educational need, V_{N_n} —permits the analysis of data generated by a three-component model of educational and training need (Misanchuk, 1982; Misanchuk & Scissons, 1978; Scissons, 1982), or, more generally, allows for the analysis of data generated by a model which incorporates four or more dimensions.

A standardized set of error weights (Figure 3) was proposed for the three-component case, and the expected marginal distributions which must be specified by the needs analyst (monotonically increasing for relevance of the task or skill to be taught and for the individual's competence at the task or skill, and normally distributed for desire to undertake further education or training in the task or skill) were identified. The statistic allows the researcher to deviate from the recommended values if there seems good reason to do so. For example, if it were part of the model of educational need that desire should count only half as heavily toward determining educational need as the other need components, the set of error weights can be adjusted to accommodate the reduced influence of desire. Or, if the researcher had evidence to show that, say, the expected marginal distribution for desire should be something other than a normal distribution, the adjustment could be made: nothing in the mathematics of determining V_{N_n} is affected by the specification of alternative error weights or expected marginal distribution schemes.

The increased sensitivity of V_{N_n} to changes in respondent distribution as compared to competing methods (Misanchuk, 1980), its ability to accommodate different numbers and relative emphases of dimensions according to user-defined models, and its relative simplicity of computation[8] argue for its application in instructional development projects, especially where large numbers of learners and/or large numbers of skills must be studied.

References

Deden-Parker, A. (1980). Needs assessment in depth: Professional training at Wells Fargo Bank. Journal of Instructional Development, 4(1), 3-9.

Gagné, R. M. (1977, April). Discovering educational goals - a research problem. Paper presented at the Annual Meeting of the Association for Educational Communications and Technology, Dallas, TX.

Hildebrand, D. K., Laing, J. D., & Rosenthal, H. (1977a). Analysis of ordinal data. Sage university paper series on quantitative applications in the social sciences, no. 8. London: Sage.

Hildebrand, D. K., Laing, J. D., & Rosenthal, H. (1977b). Prediction analysis of cross classifications. New York: Wiley.

Jones, W. A., & Sommers, P. A. (1975). Comprehensive needs assessment: An inferential approach. Educational Technology, 15(4), 54-57.

Kaufman, R. A. (1972). Educational system planning. Englewood Cliffs, NJ: Prentice-Hall.

Kaufman, R. A. (1977a). A possible taxonomy of needs assessments. Educational Technology, 17(11), 60-64.

Kaufman, R. A. (1977b). Needs assessments: Internal and external. Journal of Instructional Development, 1(1), 5-8.

Kaufman, R., & English, F. W. (1979). Needs assessment: Concept and application. Englewood Cliffs, NJ: Educational Technology Publications.

Kaufman, R., Stakenas, R. G., Wager, J. C., & Mayer, H. (1981). Relating needs assessment, program development, implementation, and evaluation. Journal of Instructional Development, 4, 17-26.

LeSage, E. C., Jr. (1980). A quantitative approach to educational needs assessment. Canadian Journal of University Continuing Education, 6(2), 6-13.

Misanchuk, E. R. (1980). A methodological note on quantitative approaches to needs assessment. Canadian Journal of University Continuing Education, 7(1), 31-33.

Misanchuk, E. R. (1982, May). Toward a multi-component model of educational and training needs. Paper presented at the Annual Meeting of the Association for Educational Communications and Technology, Dallas, TX.

Misanchuk, E. R. (in press). The analysis of multi-component educational and training needs data. Journal of Instructional Development.

Misanchuk, E. R., & Scissons, E. H. (1978, November). The Saskatoon business training needs identification study: Final report. Saskatoon, Sask.: University of Saskatchewan, Extension Division.

Monette, M. L. (1977). The concept of educational need: An analysis of selected literature. Adult Education, 27, 116-127.

Reynolds, H. T. (1977) The analysis of cross-classifications. New York: Free Press.

Rossett, A. (1982). A typology for generating needs assessments. Journal of Instructional Development, 6(1), 29-33.

Roth, J. (1977). Needs and the needs assessment process. Evaluation News, 5, 15-17.

Scissons, E. H. (1982). A typology of needs assessment definitions in adult education. Adult Education, 33, 20-28.

Scriven, M., & Roth, J. (1977). Needs assessment. Evaluation News, 2, 25-28.

Spitzer, D. R. (1981). Analyzing training needs. Educational Technology, 21(11), 36-37.

Witkin, B. R. (1976, April). Needs assessment models: A critical analysis. Paper presented at the American Educational Research Association Annual Meeting, San Francisco, CA.

Witkin, B. R. (1977). Needs assessment kits, models and tools. Educational Technology, 17(11), 5-18.

Footnotes

1. The author would like to acknowledge the assistance of R. A. Yackulic with some of the technical details of the paper. T. M. Schwen graciously read and commented upon an earlier draft of this paper.

2. We acknowledge the distinction between training and education, but for the sake of ease in reading we will, in this paper, henceforth avoid specifying each of them by treating the two as synonymous. From the point of view of needs assessment, whether the need is for training or for education is largely irrelevant, as the needs assessment procedures are virtually the same. The language used in this paper typically refers to the assessment of job-related training.

3. The terms task and skill will be used in this paper to mean approximately the same thing: a job-related activity that can be learned. Tasks or skills—as the terms are used here—include everything from specific psychomotor activities to complex groups of activities that may involve cognitive and/or psychomotor (and perhaps even affective) components. For example, while typing at 60 wpm would certainly qualify as a skill under the definition used here, so would preparing an income tax return, or counselling employees. Equally, the terms are meant to apply to such multi-faceted activities as using computers, and wage and salary administration. Hereafter, the terms skill and task will be used interchangeably.

4. This statement is true for the generic del as defined by Hildebrand et al. (1977a, 1977b). However, for the needs analysis case, the expected error rate is specified by the needs analyst as monotonically increasing along both dimensions as we move away from the upper left hand corner of the two-dimensional matrix (Misanchuk, 1983). The point will be raised again later in this paper.

5. The dot notation indicates summation over all values of the subscript which is replaced by the dot. For example, $P_{i.}$ means the proportion of i considered over all values of j; $P_{.j}$ means the proportion of j over all i.

6. Multiple dot notation is read in a manner similar to single dot notation. Hence, $P_{i..}$ is the proportion of i over all categories of j and k, $P_{.j.}$ is the proportion of j over all i and k, and so on.

7. Some experience with the problem suggests that there may be a more complex relationship between competence and desire for further training than this assumption acknowledges: It often seems that learners want to learn more about something that they already do reasonably well (the "preaching to the converted" syndrome), making the assumption of a normal distribution somewhat questionable. However, the assumption will suffice until further research can establish more accurately the exact relationship between learners' competence and their desire for further training. At such time, an accommodation can be made by simply changing the expected marginal distribution.

8. The computation of $V_{\underline{N}_3}$ is well within the scope of a hand calculator, albeit somewhat tedious when numerous skills are being studied. Existing computer programs can be acquired, or can easily be written to ease the labor.

Table 1

Values of Trivariate Del for Example Data in Figure 6

Matrix	$\nabla_{\underline{N}_3}$
(a)	0.9100
(b)	0.8621
(c)	0.8143
(d)	0.6743
(e)	0.3655
(f)	-0.1673

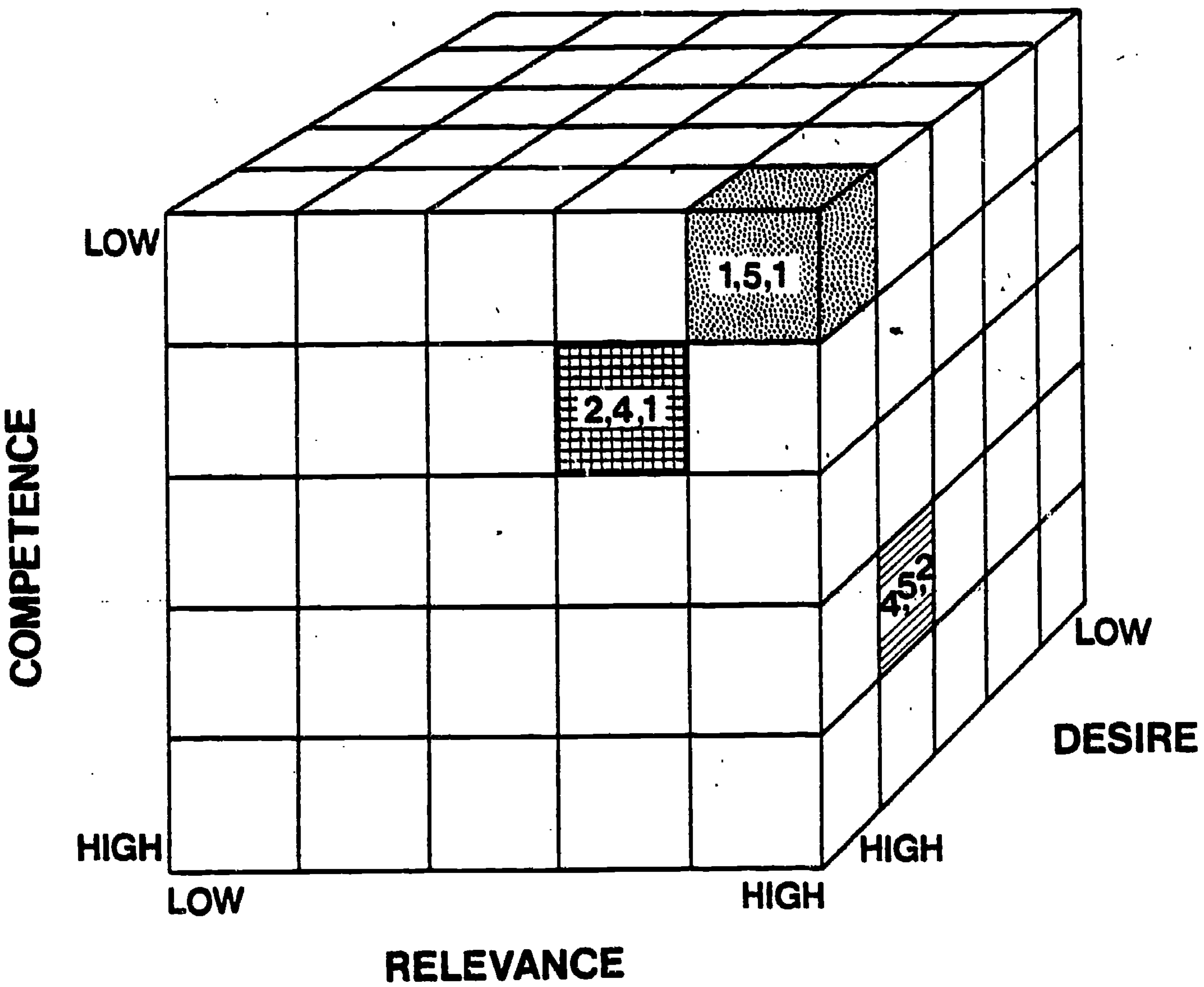

Figure 1. Representation of a three-dimensional model of educational need. Each dimension of the cube is arbitrarily divided into five categories for convenience in data-collection; some other number of categories could be used.

	x_1			x_2		
	z_1	z_2	z_3	z_1	z_2	z_3
y_1						
y_2						
y_3						

(a)

	x_1 & z_1	x_1 & z_2	x_1 & z_3	x_2 & z_1	x_2 & z_2	x_2 & z_3
y_1						
y_2						
y_3						

(b)

Figure 2. Two ways of graphically representing three-dimensional data. Once represented in the arrangement in (b), the data can be used to compute a two-dimensional V in the conventional way.

DESIRE FOR FURTHER TRAINING

	HIGH																								LOW
	RELEVANCE					RELEVANCE					RELEVANCE					RELEVANCE					RELEVANCE				
COMPETENCE	LOW				HIGH	LOW				HIGH	LOW				HIGH	LOW				HIGH	LOW				HIGH
LOW	5774	4330	2887	1443	0000	5951	4564	3227	2041	1443	6455	5204	4082	3227	2887	7217	6124	5204	4564	4330	8165	7217	6455	5951	5774
	5951	4564	3227	2041	1443	6124	4787	3536	2500	2041	6614	5401	4330	3536	3227	7360	6292	5401	4787	4564	8292	7360	6614	6124	5951
	6455	5204	4082	3227	2887	6614	5401	4330	3536	3227	7071	5951	5000	4330	4082	7773	6770	5951	5401	5204	8660	7773	7071	6614	6455
	7217	6124	5204	4564	4330	7360	6292	5401	4787	4564	7773	6770	5951	5401	5204	8416	7500	6770	6292	6124	9242	8416	7773	7360	7217
HIGH	8165	7217	6455	5951	5774	8292	7360	6614	6124	5951	8660	7773	7071	6614	6455	9242	8416	7773	7360	7217	10000	9242	8660	8292	8165

Figure 3. Proposed standard error weights for trivariate proportionate reduction in error index of educational need using five-point scales. Entries are X 10^{-4}; decimals points omitted for clarity.

DESIRE FOR FURTHER TRAINING

HIGH — LOW

	RELEVANCE					RELEVANCE					RELEVANCE					RELEVANCE					RELEVANCE					
COMPETENCE	LOW				HIGH	LOW				HIGH	LOW				HIGH	LOW				HIGH	LOW				HIGH	
LOW				3	7				2	3																15
				9	11		1		6	1			2		4					1				1		36
										4		3														7
				2															1			2				5
HIGH																										
				14	18		1		8	8			2		4		3		1	1		2		1		63

(a)

DESIRE FOR FURTHER TRAINING

HIGH — LOW

	COMPETENCE					COMPETENCE					COMPETENCE					COMPETENCE					COMPETENCE					
RELEVANCE	LOW				HIGH	LOW				HIGH	LOW				HIGH	LOW				HIGH	LOW				HIGH	
LOW																										
							1											3						2		6
												2														2
	3	9		2		2	6												1			1				24
HIGH	7	11				3	1	4				4					1									31
	10	20		2		5	8	4				6					1	3	1			1		2		63

(b)

COMPETENCE

HIGH — LOW

	RELEVANCE					RELEVANCE					RELEVANCE					RELEVANCE					RELEVANCE					
DESIRE	LOW				HIGH	LOW				HIGH	LOW				HIGH	LOW				HIGH	LOW				HIGH	
LOW							2												1							3
									1			3								1						5
																		2		4						6
															4		1		6	1				2	3	17
HIGH									2										9	11				3	7	32
							2		3			3			4		1	2	16	17				5	10	63

(c)

Figure 4. Example data for the three-dimensional case.

DESIRE FOR FURTHER TRAINING

HIGH — LOW

	COMPETENCE					COMPETENCE					COMPETENCE					COMPETENCE					COMPETENCE				
RELEVANCE	LOW				HIGH	LOW				HIGH	LOW				HIGH	LOW				HIGH	LOW				HIGH
LOW	5774	5951	6455	7217	8165	5951	6124	6614	7360	8292	6455	6614	7071	7773	8660	7217	7360	7773	8416	9242	8165	8292	8660	9242	10000
	4330	4564	5204	6124	7217	4564	4787	5401	6292	7360	5204	5401	5951	6770	7773	6124	6292	6770	7500	8416	7217	7360	7773	8416	9242
	2887	3227	4082	5204	6455	3227	3536	4330	5401	6614	4082	4330	5000	5951	7071	5204	5401	5951	6770	7773	6455	6614	7071	7773	8660
	1443	2041	3227	4564	5951	2041	2500	3536	4787	6124	3227	3536	4330	5401	6614	4564	4787	5401	6292	7360	5951	6124	6614	7360	8292
HIGH	0000	1443	2887	4330	5774	1443	2041	3227	4564	5951	2887	3227	4082	5204	6455	4330	4564	5204	6124	7217	5774	5951	6455	7217	8165

(a)

COMPETENCE

HIGH — LOW

	RELEVANCE					RELEVANCE					RELEVANCE					RELEVANCE					RELEVANCE				
DESIRE	LOW				HIGH	LOW				HIGH	LOW				HIGH	LOW				HIGH	LOW				HIGH
LOW	10000	9242	8660	8292	8165	9242	8416	7773	7360	7217	8660	7773	7071	6614	6455	8292	7360	6614	6124	5951	8165	7217	6455	5951	5774
	9242	8416	7773	7360	7217	8416	7500	6770	6292	6124	7773	6770	5951	5401	5204	7360	6292	5401	4787	4564	7217	6124	5204	4564	4330
	8660	7773	7071	6614	6455	7773	6770	5951	5401	5204	7071	5951	5000	4330	4082	6614	5401	4330	3536	3227	6455	5204	4082	3227	2887
	8292	7360	6614	6124	5951	7360	6292	5401	4787	4564	6614	5401	4330	3536	3227	6124	4787	3536	2500	2041	5951	4564	3227	2041	1443
HIGH	8165	7217	6455	5951	5774	7217	6124	5204	4564	4330	6455	5204	4082	3227	2887	5951	4564	3227	2041	1443	5774	4330	2887	1443	0000

(b)

Figure 5. Reconfigured error weights.

DESIRE FOR FURTHER TRAINING

HIGH — LOW

COMPETENCE	RELEVANCE					RELEVANCE					RELEVANCE					RELEVANCE					RELEVANCE				
	LOW				HIGH	LOW				HIGH	LOW				HIGH	LOW				HIGH	LOW				HIGH
LOW				10	180				10	10															
				10	10				10	10															
HIGH																									

(a)

			1	10	161			1	10	10			1	1	1										
			1	10	10			1	10	10			1	1	1										
			1	1	1			1	1	1			1	1	1										

(b)

			2	10	142			2	10	10			2	2	2										
			2	10	10			2	10	10			2	2	2										
			2	2	2			2	2	2			2	2	2										

(c)

(figure continued)

DESIRE FOR FURTHER TRAINING

COMPETENCE	HIGH																								LOW
	RELEVANCE					RELEVANCE					RELEVANCE					RELEVANCE					RELEVANCE				
	LOW				HIGH	LOW				HIGH	LOW				HIGH	LOW				HIGH	LOW				HIGH
LOW		1	2	10	105		1	2	10	10		1	2	2	2		1	1	1	1					
		1	2	10	10		1	2	10	10		1	2	2	2		1	1	1	1					
		1	2	2	2		1	2	2	2		1	2	2	2		1	1	1	1					
		1	1	1	1		1	1	1	1		1	1	1	1		1	1	1	1					
HIGH																									

(d)

1	1	2	10	44	1	1	2	10	10	1	1	2	2	2	1	1	1	1	1	1	1	1	1	1
1	1	2	10	10	1	1	2	10	10	1	1	2	2	2	1	1	1	1	1	1	1	1	1	1
1	1	2	2	2	1	1	2	2	2	1	1	2	2	2	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

(e)

1	2	1	10	22	1	2	1	10	10	1	2	1	1	1	1	2	2	2	2	1	1	1	1	1
1	2	1	10	10	1	2	1	10	10	1	2	1	1	1	1	2	2	2	2	1	1	1	1	1
1	2	1	1	1	1	2	1	1	1	1	2	1	1	1	1	2	2	2	2	1	1	1	1	1
1	2	2	2	2	1	2	2	2	2	1	2	2	2	2	1	2	2	2	2	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	1	1	1	1	1

(f)

Figure 6. Example data illustrating systematically decreasing educational need.